Models of Human Movement Detection in Video Frames Using Machine Learning Technology

Arwa Darwish Alzughaibi

University of Technology Sydney
Faculty of Engineering and Information Technology

Supervisor

Dr. Zenon Chaczko

This dissertation is submitted for the degree of Doctor of Philosophy

Autumn 2019

Certificate of Original Authorship

I, Arwa Alzughaibi declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature: Signature removed prior to publication.

Date: 25-10-2019

Abstract

In recent years, human movement detection has been a very active and vibrant field of research. Examples of successful applications that adopt discoveries and developments in the human movement detection domain include pedestrian detection, intelligent monitoring systems and activity recognition. Thus far, much research work has been done and many different approaches explored to optimise accuracy, performance and productivity of human detection system. However, there are still many open questions related to these issues remain unresolved. The following study aims to develop a more effective human detection system that is able to operate in various environmental conditions and application contexts including illumination changes, pose and scale differences. The existing solutions for human detection and tracking systems often do not produce reliable and consistent results without considering changes in environmental conditions. In this work, a novel illumination invariant human detection algorithm is proposed which applies an alternative approach for the selection of orientation extraction and texture extraction features to identify human shapes in various illumination and contrast invariant conditions. An innovative human detection approach is also proposed to resolve and improve results of pose invariant cases. This research work involves the exploration of feature extraction techniques that offer superior results when dealing with human subjects in pose invariant conditions. Another innovation of this investigation is the design of a human detection and tracking model that can work in situations where human subjects are occluded within frames. In the proposed models, several pre-processing and post-processing stages are used for reducing detection errors and to improve the model performance. These approaches help to classify the frame contents more efficiently. The proposed computational solutions are extensively tested and performance evaluated using the standard datasets. The resulting output is encouraging when compared to the reported and the State-Of-The-Art human detection algorithms. The newly developed methods are tested using two practical applications and are included in this thesis as action research studies. In the first action study, children activity monitoring system is built to test the human movement detection algorithms, whilst the second action study involves a construction of an expert system for counting humans in a moving crowd to validate the effectiveness of the proposal computational models. These two key action research studies report high feasibility and viability of the proposed solutions

إهداء

أهدي ثمرة جهدي

إلى من قاسمني الحياه بيسرها وعسرها

إلى من تحلو الحياه بوجودهم

إلى من هم لحياتي حياة .. إلى قرة عيني ونبض قلبي

رنيم ♥ فرح ♥ إبراهيم

Acknowledgments

In the name of Allah, the Most Gracious and the Most Merciful

Alhamdulillah, I praise and thank Allah SWT for his greatness and for giving me the strength, knowledge, ability, and courage to complete this thesis. Without his blessings, this work could not exist.

First and foremost, I offer my special appreciation to Dr. Zenon Chaczko, my research supervisor for providing his heartfelt support, inspiration, guidance and for his invaluable help of constructive comments and suggestions at all times. I would also like to thank Dr. Chris Chiu who helped to review and proofread this thesis, as well as to my fellow friends at the UTS Faculty of Engineering and Information Technology for their support throughout my studies.

On the personal front, I would like to express my sincerest gratitude to my parents, husband, daughters, son and sister Mashael, for their love, sincere prayers and encouragement throughout my study.

Then, to those who supported me throughout my journey to achieving my goal, your kindness means a lot to me. Thanks for being there when my family couldn't.

Finally, my acknowledgment would be incomplete without thanking and expressing my deepest gratitude to the Saudi Arabia Culture Mission and Taibah University for providing a chance and funding to engage this research.

Contents

List of Figures

List of Tables

List of Algorithms

Nomenclature

AHE	Adaptive Histogram Equalization
AR	Augmented Reality
AUC	Area Under the Curve
CPU	Central Processing Unit
CSLBP	Centre Symmetric LBP
DPM	Deformable Part Model
EKF	Extended Kalman Filter
FN	False Negative
FNR	False Negative Rate
FP	False Positive
FPpI	False Positive per Image
FPR	False Positive Rate
HCI	Human Computer Interaction
LBP	Local Binary Pattern
LDCF	Local Decorrelated Channel Features
LTP	Local Ternary Pattern
MIL	Multiple Instance Learning
MoG	Mixture of Gaussian
MR	Miss Rate
NRLBP	Non-Redundant Local Binary Patterns
OF	Optical Flow
PSF	Point Spread Function
ROC	Receiver Operator Characteristics
TN	True Negative
TP	True Positive
TPR	True Positive Rate

Terms and Definitions

Accuracy	Refers specifically to the closeness of the measurement to the true value.
Bounding Box	The kind of method of annotation to mark a region or object of interest from a source image. In a two-dimensional image, it is typically representative of the coordinate value of the top-left pixels, the height of the box and width of the box.
Classification	The process of image classification on the premise of recognising characteristics or features within an image.
Computer Vision	People use their brains and eyes to visualise the world that they perceive, thus computer vision is the computer science aiming to provide a similar, or optimal ability to a computer or machine. Computer vision deals with the automated extracting, analysing and understanding of meaningful data from an image or image sequence. It deals with the design of an algorithmic and theoretical construct to accomplish the automatic understanding of visuals.
Confidence Score	Defining the event probability, or input probability to fall in differing classes. If a class has a great probability then confidence is high. The value of confidence is calculated for one input, and providing meaning as to the algorithm's confidence for the class.
Detection	The process of determining and finding notable objects in an image.
Feature Extraction	The process of detecting and extracting the most representing data from a raw source as features, then deriving discovered features from the original such as to lower the cost of measuring the features, improve efficiency of the classifier and enable classification with higher accuracy.
Feature Representation	The process of providing a uniquely representative point for every sequence of video based on features that are extracted.
Ground Truth	The term used to describe the data offered by observing directly, that is empirical evidence, instead of the information made by inferences.

Human Detection	The task of localisation of human instances that are present within an image, and it is mostly done by finding all the locations in the image, for all available scales, and making a comparison in a small region at every location with well-known people patterns or templates.
Precision	Refers specifically to the closeness of the measurement to each other.
Receiver Operating Characteristic Curve	It is a curve represented as a graphic plot chart, depicting the diagnosis of a binary classifier system, because of the variation of its discriminatory threshold.